PROCEEDINGS OF TDWG

Biodiversity Information Science and Standards 1: e20435
doi: 10.3897/tdwgproceedings.1.20435

OPEN ACCESS

Conference Abstract

Semantically Defining Populations for 'Omics Research

Ramona L. Walls‡, Pier Luigi Buttigieg§

‡ CyVerse, Tucson, United States of America
§ Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Corresponding author: Ramona L. Walls (rlwalls2008@gmail.com)

Received: 18 Aug 2017 | Published: 20 Aug 2017

Citation: Walls R, Buttigieg P (2017) Semantically Defining Populations for 'Omics Research. Proceedings of TDWG 1: e20435. https://doi.org/10.3897/tdwgproceedings.1.20435

Abstract

The study of populations is central to ‘omics research, whether sequencing environmental samples, controlling for population structure when looking for genetic variation within a species, or studying the evolution of large clades. Researchers use different operational definitions of populations and communities, via the highly varied creation of operational taxonomic units (OTUs) and, in some cases, use of unclustered sequences. The use of different methods, even within one study type (Swarm, UCLUST, CD HIT, etc.), creates very different OTUs, possibly affecting interpretation and leading to questionable reproducibility. The Population and Community Ontology (PCO) offers the semantics to clarify exactly which collection of organisms (i.e., ecological community or population) was used in an investigation. When combined with methods for standardizing observational data from the BioCollections Ontology (BCO), protocol classes from the Ontology for Biomedical investigations (OBI), and characterization of environments from the Environment Ontology (ENVO), PCO can fully describe the methods used to derive organismal or species-based (i.e. taxonomic) OTUs used for biodiversity analysis and monitoring. PCO is not well suited to describe “OTUs” based on sequence variants that may or may not map to population or individual level variation (e.g., output of some clustering algorithms). In this case, the Sequence Ontology (SO) may be more appropriate. This presentation will describe the key ontology design patterns used in the PCO and provide examples of how and when PCO and related ontologies should be used in omics research, with a focus on environmental/metagenomic sequencing applications.

© Walls R, Buttigieg P. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords

ontology, population, ecological community, omics, OTU, semantics, sequence

Presenting author

Ramona L. Walls